

МРНТИ 14.01.75

<https://doi.org/10.26577/JES872202611>Л.Е. Шинетова¹, Г.К. Жаббаева¹, Ж.Ж. Бейсенова^{1*}Е.В. Адулова², С.С. Есеналиева³,¹Национальный центр тестирования, Астана, Казахстан²«Центр педагогических измерений» АОО Назарбаев Интеллектуальные школы, Астана, Казахстан³Национальный центр повышения квалификации «Өрлеу», Астана, Казахстан

*e-mail: zhanar-beisenova@mail.ru

ПСИХОМЕТРИЧЕСКИЙ АНАЛИЗ УСТОЙЧИВОСТИ ЕНТ-2025: ВЛИЯНИЕ ТЕРРИТОРИАЛЬНЫХ И ИНСТИТУЦИОНАЛЬНЫХ ФАКТОРОВ

В статье проводится глубокое исследование психометрических свойств Единого национально-го тестирования (ЕНТ) 2025 года, выступающего ключевым инструментом академического отбора в Казахстане. Актуальность работы обусловлена необходимостью обеспечения образовательного равенства и объективности оценивания в условиях реформирования системы высшего образования. Цель исследования заключается в верификации устойчивости тестовых заданий и анализе того, насколько территориальные и институциональные характеристики, такие как регион проживания и тип образовательного учреждения, определяют итоговый результат абитуриента. В качестве методологической базы использован синтез классической теории тестов (СТТ) и современной теории (ИРТ). Особое внимание уделено DIF анализу, позволяющему выявить скрытую предвзятость тестов по отношению к различным группам респондентов. Результаты анализа подтверждают наличие устойчивой корреляции между социально-географическим положением учащегося и его академическими достижениями, фиксируя статистически значимый разрыв в качестве подготовки между городом и селом. Авторы пришли к выводу, что, несмотря на высокую внутреннюю валидность тестов, институциональные факторы продолжают играть роль «невидимых барьеров» в доступе к высшему образованию. Практическая значимость исследования состоит в возможности интеграции полученных данных в стратегию развития Национального центра тестирования для выравнивания образовательных возможностей в масштабах всей страны.

Ключевые слова: ЕНТ, психометрический анализ, академическое неравенство, ИРТ-моделирование, DIF.

Sh.B. Altybayeva¹, K.T. Turalbaeva¹, L.E. Shinetova¹,
Y.V. Adulova², S.S. Yessenaliyeva³¹National Testing Center, Astana, Kazakhstan²«Center for Pedagogical Measurements» under AEO «Nazarbayev Intellectual Schools», Astana, Kazakhstan³National Center for Professional Development «Orleu», Astana, Kazakhstan

*e-mail: zhanar-beisenova@mail.ru

Psychometric Stability Analysis of UNT-2025: Impact of Territorial and Institutional Factors

This article presents an in-depth study of the psychometric properties of the 2025 Unified National Testing (UNT), which serves as a cornerstone of academic selection in Kazakhstan. The relevance of this work is driven by the imperative to ensure educational equity and assessment objectivity amidst the ongoing reforms in the higher education system. The research objective is to verify the stability of test items and analyze the extent to which territorial and institutional characteristics, such as region of residence and school type, determine an applicant's final performance. A synthesis of Classical Test Theory (CTT) and Item Response Theory (IRT) forms the methodological framework. Special emphasis is placed on DIF analysis, which allows for the identification of latent test bias across various respondent groups. The findings confirm a steady correlation between a student's socio-geographic location and their academic achievements, highlighting a statistically significant gap in the quality of preparation between urban and rural areas. The authors conclude that while the tests exhibit high internal validity, institutional factors continue to act as «invisible barriers» to higher education access. The practical significance of this study lies in the potential integration of the findings into the National Testing Center's development strategy to level educational opportunities across the nation.

Keywords: UNT, psychometric analysis, academic inequality, IRT modeling, DIF.

Ш.Б. Алтыбаева¹, К.Т. Туралбаева¹, А.Е. Шинетова¹,
Е.В. Адулова³, С.С. Есеналиева³

¹Ұлттық тестілеу орталығы, Астана, Қазақстан

²«Педагогикалық өлшеулер орталығы» Назарбаев Зияткерлік мектептері ДББҰ, Астана, Қазақстан

³«Өрлеу» біліктілікті арттыру ұлттық орталығы, Астана, Қазақстан

*e-mail: zhanar-beisenova@mail.ru

ҰБТ-2025 тұрақтылығын психометрикалық талдау: аумақтық және институционалдық факторлардың әсері

Мақалада Қазақстандағы академиялық іріктеудің негізгі құралы болып табылатын 2025 жылғы Ұлттық бірыңғай тестілеудің (ҰБТ) психометрикалық қасиеттеріне терең зерттеу жүргізілген. Жұмыстың өзектілігі жоғары білім беру жүйесін реформалау жағдайында білім беру теңдігі мен бағалаудың объективтілігін қамтамасыз ету қажеттілігінен туындайды. Зерттеудің мақсаты – талапкердің қорытынды нәтижесін тұрғылықты жері мен оқу орнының түрі сияқты аумақтық және институционалдық сипаттамалардың тест тапсырмаларының тұрақтылығын тексеру және талдау болып табылады. Әдістемелік негіз ретінде классикалық тест теориясы (СТТ) мен заманауи тест теориясының (IRT) синтезі қолданылды. Респонденттердің әртүрлі топтарына қатысты тесттердегі жасырын ауытқуларды анықтауға мүмкіндік беретін DIF талдауына ерекше назар аударылды. Талдау нәтижелері қала мен ауыл арасындағы дайындық сапасының статистикалық маңызды алшақтығын тіркей отырып, студенттің әлеуметтік-географиялық жағдайы мен оның оқу жетістіктері арасында тұрақты корреляцияның болуын растайды. Авторлар тесттердің жоғары ішкі валидтілігіне қарамастан, институционалдық факторлар жоғары білімге қолжетімділікте «көрінбейтін кедергілер» рөлін атқаруды жалғастыруда деген қорытындыға келеді. Зерттеудің тәжірибелік маңыздылығы алынған деректерді бүкіл ел бойынша білім беру мүмкіндіктерін теңестіру үшін Ұлттық тестілеу орталығының даму стратегиясына біріктіру мүмкіндігі болып табылады.

Түйін сөздер: ҰБТ, психометрикалық талдау, академиялық теңсіздік, IRT-моделдеу, DIF.

Введение

Академическая успеваемость традиционно выступает ключевым индикатором качества системы образования, отражая как индивидуальные достижения обучающихся, так и эффективность национальной образовательной политики. В связи с этим стандартизированные тесты, оценивающие предметные знания, когнитивные навыки и уровень подготовленности, на протяжении десятилетий используются в мировой практике для академического отбора и прогнозирования будущей успешности. При этом многочисленные исследования подтверждают, что психометрический анализ результатов тестирования, а также целостность и качество таких измерений являются необходимыми условиями получения обоснованных и достоверных научных выводов (Barry et al., 2014; Bull et al., 2019).

В Казахстане одним из наиболее значимых экзаменов является Единое национальное тестирование (далее – ЕНТ), которое выполняет функции вступительного экзамена в высшие учебные заведения и механизма распределения государственных грантов. По своей сути ЕНТ выступает инструментом вертикальной академической мобильности, обеспечивая прозрачность и сопоставимость достижений выпускников в масштабах страны.

В рамках выполняемых функций в последние годы НЦТ реализует стратегию повышения психометрического качества тестовых материалов, ориентируясь на международные стандарты образовательного измерения (AERA et al., 2014; Abdrasilov et al., 2025). Значимость и масштаб проводимой работы подтверждаются статистическими данными: в 2025 году в ЕНТ приняли участие 728 983 человека. В этих условиях обеспечение надежности, объективности и валидности теста приобретает не только методологический, но и социальный характер, поскольку результаты определяют доступ к высшему образованию.

Эволюция ЕНТ стала ответом на системные вызовы, включая коррупцию и недобросовестные практики ранних лет, о чем пишет Jumabayeva (2016). В рамках реформ внедрен многоформатный тест с блоками по читательской и математической грамотности, что ознаменовало переход от проверки запоминания фактов к оценке навыков анализа и применения информации (egov.kz, 2024; Didarbekova et al., 2025). Современная структура включает три обязательных компонента («История Казахстана», «Грамотность чтения», «Математическая грамотность») и два профильных предмета по выбору (Национальный центр тестирования, 2025).

Ключевым нововведением стала возможность многократной сдачи теста в течение года, что снижает психологическую нагрузку и влияние случайных. Глубокая трансформация произошла в 2021 году с переводом экзамена в электронный формат, усилившим академическую честность за счет систем прокторинга, видеонаблюдения и биометрической идентификации. Однако такая гибкость системы ставит перед психометриками задачу оценки эквивалентности и устойчивости результатов между попытками.

Несмотря на реформы, сохраняются вызовы, в частности «обратный эффект» (washback effect), когда ориентация на тест сужает содержание обучения, усиливает роль натаскивания и усугубляет неравенство между городскими и сельскими школами (Mingisheva, 2023; Chankseliani et al., 2020; OECD, 2020). Это противоречит цели ЕНТ – обеспечить справедливую и объективную оценку знаний.

Постановка проблемы. Несмотря на модернизацию ЕНТ, остаются вопросы психометрических характеристик, влияющих на надежность, валидность и справедливость экзамена. Особое значение имеют территориальные различия в доступе к ресурсам, педагогическому составу и инфраструктуре, которые могут влиять на вариативность результатов и стабильность показателей между попытками (OECD, 2020; Mingisheva, 2023; Chankseliani et al., 2020). Учитывая high-stakes характер ЕНТ, отсутствие стабильности может усиливать неравенства и снижать доверие к системе.

В связи с этим необходим психометрический анализ результатов ЕНТ 2025 года с учетом территориальных факторов для оценки сопоставимости показателей и выявления системных различий. Актуальность настоящего исследования обусловлена необходимостью понимания влияния внешних переменных, таких как регион обучения, тип населенного пункта и местоположение школы, на динамику баллов и стабильность образовательных достижений (Koreyeva, 2025).

Цель исследования – комплексный психометрический анализ результатов ЕНТ 2025 года с акцентом на:

- оценку устойчивости результатов между первой и второй попытками;
- изучение влияния региона обучения, типа населенного пункта и местоположения школы (город/село) на стабильность и сопоставимость показателей.

Обзор литературы

Психометрическое качество крупномасштабных тестов, таких как ЕНТ, определяется тремя взаимосвязанными конструктами: надежностью (точностью и устойчивостью измерений), валидностью (обоснованностью интерпретаций результатов) и управляемостью (практичностью, эффективностью и справедливостью применения) (AERA et al., 2014).

Messick (1994) предложил концепцию единой валидности, где она рассматривается как интегральная оценка обоснованности интерпретаций тестовых баллов.

Для экзаменов с высокими ставками (high-stakes exams) вроде ЕНТ ключевой является прогностическая валидность – способность предсказывать академический успех в вузе. Отсутствие такой валидности делает тест методологически непригодным для отбора (Marini et al., 2020; Sanchez, 2025).

Обеспечение валидности подразумевает контроль на всех этапах: от разработки заданий до анализа результатов. Угрозы включают нерепрезентативную выборку, некорректные формулировки и нарушения в обработке данных, что может привести к несправедливым интерпретациям. Содержательная валидность отражает соответствие теста учебной программе, а конструктивная – способность измерять целевые компетенции (AERA et al., 2014). В анализе ЕНТ важно учитывать угрозы конструктивной валидности: конструктивную недопредставленность (underrepresentation), постороннюю сложность (construct-irrelevant variance) и постороннюю легкость, чтобы тест оценивал реальные навыки, а не внешние факторы, эти аспекты подробно описал Messick (1995). Аналогии с SAT и ACT показывают, что такие тесты измеряют приобретенные когнитивные навыки, а не врожденные способности (Amrein-Beardsley et al., 2025).

Контроль предсказуемости содержания теста важен для предотвращения сужения учебной программы и «натаскивания», на что указывает Messick (1996). Messick подчеркивает, что валидность не оценивается напрямую через washback effect из-за его сложности; вместо этого анализируются свойства заданий (аутентичность, прямота) в контексте конструктивной валидности.

Надежность – устойчивость результатов – является предпосылкой валидности: ненадежный тест не может быть валидным, как утверждает Smith (2001). Метрики, такие как коэффициент

Кронбаха (α) и композитная надежность (CR), оценивают внутреннюю согласованность, а низкая надежность повышает стандартную ошибку измерения (SEM), рискуя несправедливостью (Algina & Penfield, 2009; Woodruff & Wu, 2012). Управляемость касается практической реализации теста для стейкхолдеров (AERA et al., 2014; Smith, 2001). Dorans (1989) отмечает, что факторы включают логику, оборудование и устранение bias через анализ дифференциального функционирования заданий (DIF) (AERA et al., 2014). В контексте ЕНТ, где данные выявляют территориальные различия, DIF-анализ необходим для выявления систематических ошибок по регионам или типам поселений.

Предполагается, что территориальные факторы значимо влияют на надежность и валидность, с более низкой устойчивостью результатов в сельских районах из-за неравенства инфраструктуры, что может проявляться в DIF и снижать справедливость, как утверждает Dorans (1989). Эта теоретическая основа служит фундаментом для методологии аудита ЕНТ, включая анализ устойчивости (t-тесты, корреляции) и сопоставимости (групповые сравнения).

Настоящее исследование направлено на комплексный психометрический анализ результатов ЕНТ-2025 с целью оценки устойчивости и сопоставимости показателей между первой и второй попытками тестирования. В фокусе исследования – выявление влияния территориальных факторов, включая регион обучения, тип населенного пункта и местоположение школы (город/село), на стабильность результатов.

Проблема «разрыва между городом и селом» в образовательных достижениях Казахстана остается одной из самых острых. Результаты международных сопоставительных исследований, таких как PISA-2022, подтверждают, что учащиеся городских школ значительно опережают сверстников из сельской местности по математике и чтению (OECD, 2023). Согласно исследованию Nurbayev (2021) факторы, обуславливающие этот разрыв, включают различия в инфраструктуре, квалификации педагогических кадров, доступе к информационно-коммуникационным технологиям и социально-экономическом статусе семей. Региональные диспропорции также выражены ярко: западные и южные регионы страны традиционно показывают более низкие результаты по сравнению с городами республиканского значения и северными областями (Koreyeva, 2025). Влияние этих факторов

на стабильность результатов между попытками ЕНТ ранее не подвергалось детальному систематическому анализу в масштабе всей страны.

В рамках ЕНТ 2025 года математика представлена двумя ключевыми блоками:

- обязательный компонент «Математическая грамотность», оценивающий базовые навыки применения математических концепций в повседневных ситуациях;

- профильный предмет «Математика», фокусирующийся на углубленных знаниях и решении сложных задач.

Эти блоки играют центральную роль в общей структуре экзамена, отражая не только предметные компетенции, но и когнитивные способности абитуриентов, что согласуется с международными стандартами оценки (AERA et al., 2014).

Конструкция теста и разработка заданий осуществляется на основании Государственный образовательный стандарт (далее – ГОСО) и типовые учебные программы (далее – ТУП), включая многоуровневую экспертизу и апробацию, что обеспечивает содержательную валидность и соответствие целям оценивания (Министерство просвещения РК, 2022; Министерство образования и науки РК, 2012; 2017).

Разделение математического блока на два компонента отражает глобальный тренд на оценивание функциональной грамотности. Математическая грамотность, согласно определениям PISA, подразумевает способность формулировать, применять и интерпретировать математику в разнообразных контекстах реальной жизни (Kappassova et al., 2025).

В отличие от профильной математики, фокусирующейся на глубинных академических знаниях и решении сложных абстрактных задач, математическая грамотность оценивает базовые навыки, необходимые для эффективного функционирования в современном обществе. Исследования показывают, что учащиеся часто демонстрируют высокую результативность в процедурных задачах, но испытывают трудности при интерпретации контекстных проблем, что делает анализ устойчивости именно этого компонента особенно важным (Kappassova et al., 2025).

Материалы и методы исследования

Дизайн и выборка исследования. Исследование представляет собой лонгитюдный пси-

хотметрический аудит вторичных данных ЕНТ-2025. Цель – оценка устойчивости (test-retest reliability) и сопоставимости результатов между первой (январь) и второй (июнь) попытками, а также влияние территориальных и институциональных факторов на динамику баллов. Выборка сформирована Национальным центром тестирования (далее – НЦТ) на принципе наличия полных данных по двум попыткам. Критерии включения: участие в обеих попытках ЕНТ-2025, наличие данных по математическим дисциплинам и мета-переменным (регион, тип школы, местоположение). Итоговый размер выборки составил $N = 25\,574$ человека.

Важное ограничение выборки. Распределение по регионам (таблица 1) реплицирует демографическую структуру страны (напр., высокая доля Туркестанской области), но не является стратифицированной случайной выборкой. Это позволяет проводить сравнительный анализ внутри подгрупп, но требует осторожности при экстраполяции абсолютных значений средних баллов на всю популяцию.

Переменные.

Зависимые переменные: баллы по «Математической грамотности» (MathLiteracy1, MathLiteracy2) и профильной «Математике» (Mathematics1, Mathematics2) за первую и вторую попытки (шкала 0-10 и 0-40 соответственно).

Независимые (группирующие) переменные:

SchoolLocation: городская (1) / сельская (2) местность.

SettlementType: тип населенного пункта (А-Е, от мегаполисов до сел).

RegionCode: код региона (21 категория).

SchoolType: тип организации образования (гимназия, лицей, СОШ и др.).

Методы статистического и психометрического анализа. Для решения исследовательских задач был применен следующий аналитический конвейер:

Дескриптивный анализ: Расчет M (среднего), SD (стандартного отклонения), частот для описания структуры данных.

Оценка устойчивости (надежности повторного тестирования)

Коэффициент корреляции Пирсона (r) между баллами попыток 1 и 2. Интерпретация: $r > 0,7$ – удовлетворительная надежность для high-stakes теста (AERA et al., 2014).

Парный t -тест для оценки статистической значимости прироста баллов (эффект практики).

Расчет стандартной ошибки измерения (SEM) на основе SD и надежности для оценки доверительного интервала индивидуального балла.

Оценка влияния группирующих факторов

Однофакторный дисперсионный анализ (ANOVA) для проверки гипотезы о различии средних баллов между группами. При значимом результате F -критерия применялся пост-хок тест Шеффе для попарных сравнений.

Для визуализации и анализа силы эффекта (effect size) рассчитывался η^2 (эта-квадрат).

Косвенный анализ справедливости. Выявление паттернов, указывающих на потенциальную угрозу валидности. Например, нулевой прирост баллов у сельских учащихся по грамотности при наличии прироста у городских интерпретируется как сигнал о дифференциальном эффекте подготовки, требующем последующего углубленного DIF-анализа.

Программное обеспечение и этические аспекты. Анализ проводился в среде R (v.4.3.2) с использованием пакетов psych, lavaan, ggplot2 и dplyr. Для проведения первичного дескриптивного анализа и верификации данных на этапе ввода дополнительно применялся пакет SPSS 29.

Соблюдение этических стандартов являлось приоритетом на всех этапах работы. Нами были использованы деперсонализированные и агрегированные данные, предоставленные НЦТ, что полностью исключает доступ к персональной информации поступающих. Работа носит прикладной характер и выполнена в рамках служебного задания сотрудников НЦТ, что гарантирует легитимность и практическую значимость полученных результатов.

Ограничения методологии. Отсутствие рандомизации и контрольной группы не позволяет установить причинно-следственные связи. Выборка включает только тех, кто сдавал обе попытки, что может создавать смещение в сторону более мотивированных поступающих. В рамках данного макроанализа не оценивались психометрические параметры на уровне заданий (дискриминативность, сложность, DIF), что является задачей для последующих углубленных исследований. В модели не учитывался индивидуальный социально-экономический статус учащихся, который является мощным предиктором образовательных достижений. Его эффект частично опосредован через включенные переменные (тип школы, регион).

Результаты исследования

Анализ дескриптивной статистики (таблица 1) демонстрирует распределение единиц наблюдения ($N = 25\,574$) по регионам Республики Казахстан. Распределение является существенно неравномерным. Наиболее представительным является регион с кодом 13 – Туркестанская область (14,0% от общей выборки). Также значительный вклад в общий объем данных вносят Алматинская область (код 5, 7,7%), Ак-

тубинская область (код 4, 7,3%), город Алматы (код 2, 7,2%) и Мангистауская область (код 12, 6,3%).

Наименьшее количество наблюдений зафиксировано в Северо-Казахстанской области (код 15, 1,2%), Костанайской области (код 10, 1,7%), области Улытау (код 20, 1,6%). Таким образом, выборка преимущественно репрезентирует несколько ключевых регионов, что необходимо учитывать при дальнейшей интерпретации результатов исследования.

Таблица 1

Распределение выборки по регионам (коды 1-21)

Коды регионов (N)	Количество	Действительный процент	Накопленный процент
01	1320	5,2	5,2
02	1853	7,2	12,4
03	766	3,0	15,4
04	1871	7,3	22,7
05	1963	7,7	30,4
06	928	3,6	34
07	1246	4,9	38,9
08	1492	5,8	44,7
09	1245	4,9	49,6
10	432	1,7	51,3
11	1468	5,7	57
12	1617	6,3	63,3
13	3590	14	77,4
14	758	3,0	80,4
15	314	1,2	81,6
16	542	2,1	83,7
17	1510	5,9	89,6
18	750	2,9	92,5
19	1108	4,3	96,9
20	420	1,6	98,5
21	381	1,5	100
ИТОГО	25574	100	

Примечание: Коды регионов соответствуют классификации НЦТ. Наибольшая представленность Туркестанской области (14%) отражает демографическую структуру

Данная таблица составлена авторами на основе анализа материалов

Общая динамика результатов и надежность измерений. Анализ агрегированных данных по всей выборке ($N = 25\,574$) демонстрирует наличие положительной динамики

средних баллов между первой и второй попытками тестирования во всех блоках, что характерно для систем с возможностью передачи.

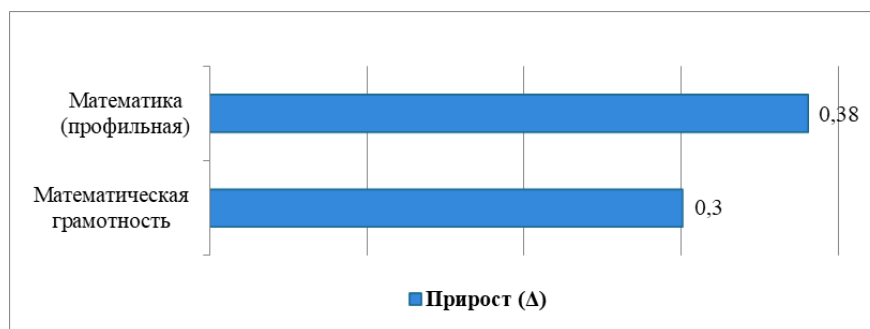
Таблица 2Динамика средних баллов и показатели надежности между первой и второй попытками тестирования ($N = 25\,574$)

Переменная	Mean (Попытка 1)	Mean (Попытка 2)	Std. Deviation 1	Std. Deviation 2	Корреляция Пирсона (r) между попытками	p-value (2-tailed)
Математическая грамотность	6,18	6,21	2,354	2,385	0,619	< 0,001
Математика (профильная)	22,70	23,08	10,866	11,086	0,881	< 0,001

Примечание: M – среднее арифметическое, SD – стандартное отклонение, r – коэффициент корреляции Пирсона
 Данная таблица составлена авторами на основе анализа материалов

Рисунок 1

Прирост показателей надежности между первой и второй попытками тестирования ($N = 25\,574$)



Источник / Примечание: Данный рисунок составлен авторами на основе анализа материалов

Результаты парного t -теста подтвердили статистическую значимость прироста баллов как по профильной математике ($p < 0,001$), так и по математической грамотности ($p = 0,018$). Однако величина прироста (effect size) существенно различается: минимальный прирост по математической грамотности ($\Delta M = +0,03$) контрастирует с более выраженным приростом по профильной математике ($\Delta M = +0,38$). Эта дивергенция указывает на различную психометрическую природу измеряемых конструктов. Профильная математика, будучи в большей степени ориентированной на академические знания и алгоритмические навыки, демонстрирует выраженный эффект обучения (practice effect) между попытками. В то же время математическая грамотность, измеряющая сформированность функциональных компетенций для решения контекстных задач, характеризуется более высокой устойчивостью (stability) результатов, что согласуется с теоретическими представлениями

о долгосрочном характере формирования подобных навыков, которую описывал Messick (1994).

Корреляционный анализ, выступающий индикатором надежности повторного тестирования (test-retest reliability), также выявил различия между конструктами. Коэффициент корреляции между попытками для профильной математики составил $r = 0,881$, что свидетельствует об отличной надежности измерительной процедуры (AERA et al., 2014). Для математической грамотности надежность оказалась умеренной ($r = 0,619$). Снижение показателя может быть объяснено сочетанием методологических факторов: относительно малым количеством заданий в субтесте (ограничение диапазона шкалы 0-10 баллов), что статистически снижает потенциальный максимум коэффициента надежности, а также большей подверженностью заданий на функциональную грамотность влиянию контекстуальных и ситуационных переменных, увеличивающих долю случайной ошибки измерения.

Влияние местоположения школы (город/село). Анализ в разрезе «город/село» выявляет устойчивое преимущество городских учащихся, которое сохраняется и даже несколько усиливается при повторном тестировании.

Сравнительный анализ результатов ЕНТ по математическим дисциплинам в разрезе типа образовательной организации выявил устойчивые и статистически значимые разли-

чия между городскими и сельскими школами. Уже по итогам первой попытки фиксируется разрыв в пользу городских учащихся как по профильной математике (24,12 против 20,90 балла), так и по математической грамотности (6,40 против 5,88 балла), что согласуется с выводами международных исследований и указывает на структурное неравенство образовательных условий.

Таблица 3

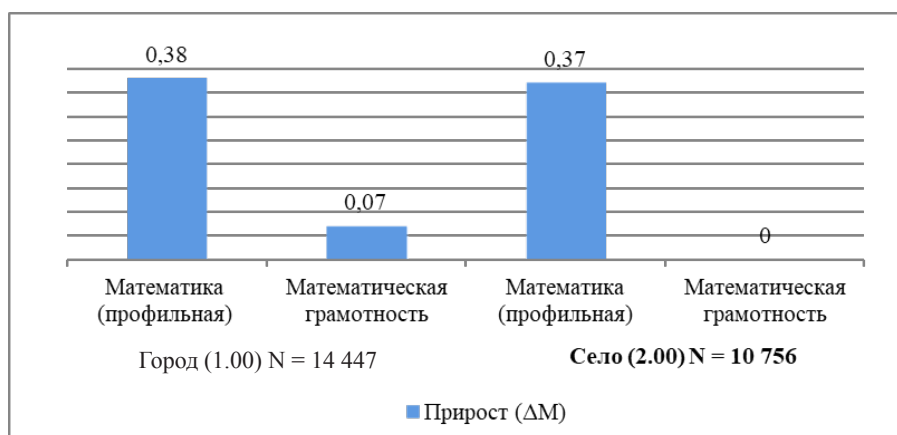
Сравнение образовательных результатов и их динамики в разрезе местоположения школы (город/село)

Местоположение	Переменная	М (Попытка 1)	М (Попытка 2)	p-value (парный t-тест)
Город (1.00) N = 14 447	Математика (профильная)	24,12	24,50	< 0,001
	Математическая грамотность	6,40	6,47	< 0,001
Село (2.00) N = 10 756	Математика (профильная)	20,90	21,27	< 0,001
	Математическая грамотность	5,88	5,88	0,748

Источник / Примечание: Данная таблица составлена авторами на основе анализа материалов

Рисунок 2

Прирост образовательных результатов в разрезе местоположения школы (город/село)



Источник / Примечание: Данный рисунок составлен авторами на основе анализа материалов

Анализ динамики между попытками демонстрирует дифференцированный эффект пересдачи. По профильной математике наблюдается сопоставимый и статистически значимый прирост результатов у учащихся городских и сельских

школ ($\approx +0,38$ балла; $p < 0,001$), что свидетельствует об универсальном эффекте повторного тестирования и дополнительной подготовки. В то же время по математической грамотности положительная динамика выявлена лишь у го-

родских учащихся (+0,07; $p < 0,001$), тогда как в сельской группе прирост отсутствует и статистически незначим.

Полученные результаты указывают на различную чувствительность оцениваемых конструкторов к эффекту практики и подтверждают, что математическая грамотность выступает индикатором устойчивого территориального неравенства в качестве образовательной среды. Отсутствие положительной динамики

в сельских школах свидетельствует об ограниченной эффективности стандартных форм подготовки и обосновывает необходимость целенаправленных интервенций, ориентированных на развитие функциональной грамотности.

Анализ по типам населенных пунктов. Классификация по типам поселений (А-Е) позволяет детализировать влияние урбанизации на образовательные результаты.

Таблица 4

Влияние типа населенного пункта на результаты ЕНТ

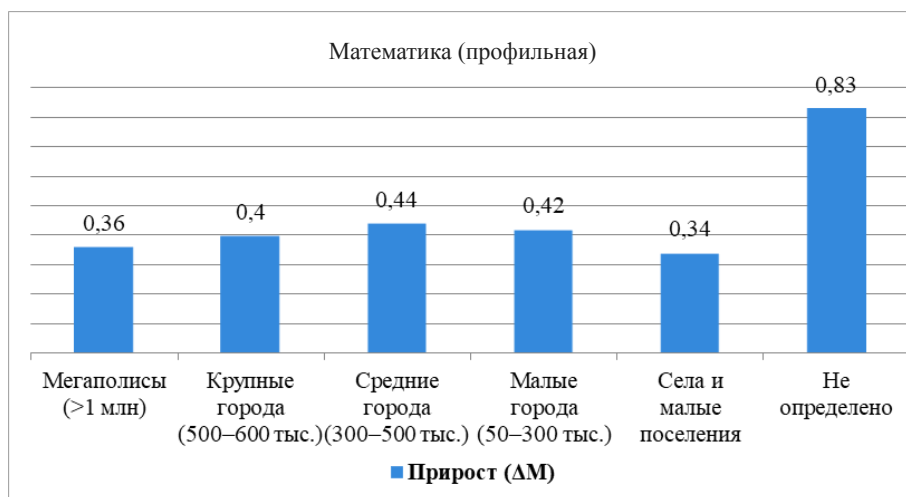
Тип поселения	Описание	Математика (профильная)		Математическая грамотность	
		М (Попытка 1)	М (Попытка 2)	М (Попытка 1)	М (Попытка 2)
A	Мегаполисы (>1 млн)	24,86	25,22	6,49	6,57
B	Крупные города (500-600 тыс.)	22,54	22,94	6,10	6,26
C	Средние города (300-500 тыс.)	25,44	25,88	6,70	6,76
D	Малые города (50-300 тыс.)	23,82	24,24	6,37	6,40
E	Села и малые поселения	21,12	21,46	5,92	5,91
F	Не определено	19,42	20,25	6,22	6,04

Примечания: М – среднее арифметическое, прирост (ΔМ) рассчитан как разность между средним значением второй и первой попытки.

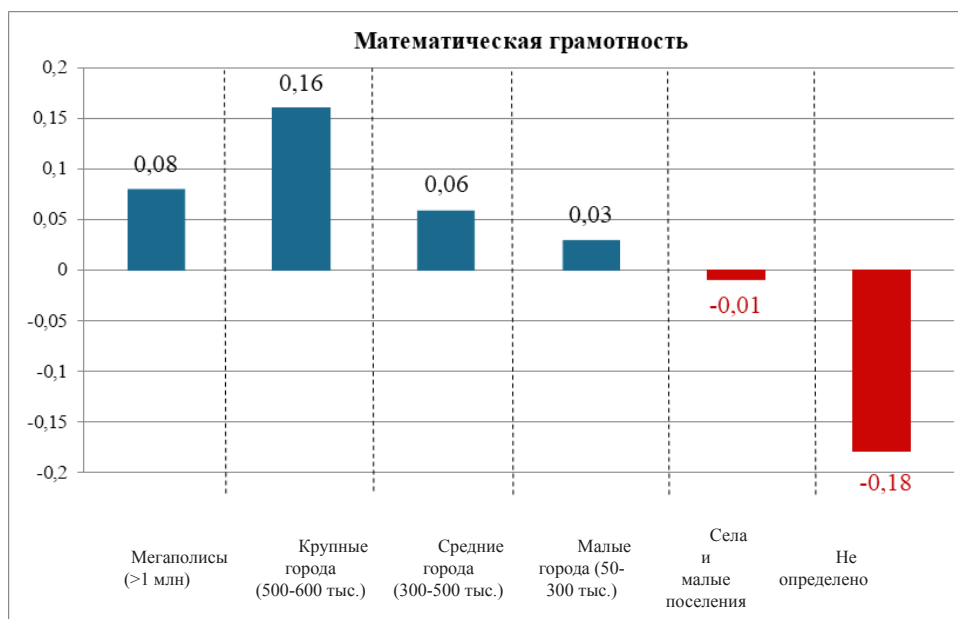
Источник / Примечание: Данная таблица составлена авторами на основе анализа материалов

Рисунок 3

Прирост по результатам «Математики (профильная)» ЕНТ в разрезе населенных пунктов



Источник / Примечание: Данный рисунок составлен авторами на основе анализа материалов

Рисунок 4*Прирост по результатам «Математической грамотности» ЕНТ в разрезе населенных пунктов**Источник / Примечание: Данный рисунок составлен авторами на основе анализа материалов*

Интересным эмпирическим результатом является тот факт, что наивысшие показатели по обоим математическим дисциплинам демонстрируют учащиеся из средних городов, опережая не только малые города и сельские населенные пункты, но и мегаполисы. По профильной математике учащиеся данной категории показывают максимальные средние баллы как в первой, так и во второй попытке (25,44 и 25,88 соответственно), а также наибольший прирост между попытками (+0,44). Аналогичная картина наблюдается и по математической грамотности (6,70 и 6,76). Данный эффект может быть связан с более высокой концентрацией специализированных школ и относительно меньшей социальной стратификацией образовательной среды по сравнению с городами-миллионниками.

Сельские и малые населенные пункты, составляющие почти половину выборки (47,8 %), демонстрируют наименьшие абсолютные результаты по обоим показателям, что подтверждает наличие системного территориального разрыва в образовательных достижениях. При этом динамика между попытками носит асимметричный характер: по профильной математике

фиксируется статистически значимый положительный прирост (+0,34), тогда как по математической грамотности прогресс отсутствует (5,92 против 5,91). Это дополнительно подтверждает вывод о том, что функциональные, контекстно-зависимые навыки в условиях сельской образовательной среды в краткосрочной перспективе не поддаются коррекции стандартными механизмами подготовки.

В совокупности полученные результаты указывают, что территориальный фактор влияет не только на уровень математических достижений, но и на характер их динамики. При этом профильная математика демонстрирует высокую чувствительность к эффекту повторного тестирования, а математическая грамотность выступает индикатором различий образовательной среды, что подчеркивает необходимость мер, направленных на поддержку формирования функциональных математических навыков.

Влияние типа школы и образовательной программы. Тип образовательного учреждения выступает мощным предиктором результатов, отражая селективность и качество преподавания.

Таблица 5

Сравнение образовательных результатов в разрезе типа школы

Тип школы	Описание	Переменная	М (Попытка 1)	М (Попытка 2)	p-value (парный t-тест)
1	Школа-гимназия N= 3258	Математика (профильная)	23,87	24,25	< 0,001
		Математическая грамотность	6,36	6,42	0,111
2	Школа-лицей N= 2977	Математика (профильная)	25,64	25,88	0,014
		Математическая грамотность	6,63	6,72	0,018
4	Средняя школа N= 17128	Математика (профильная)	21,39	21,75	< 0,001
		Математическая грамотность	5,96	5,98	0,432
5	Empty value N=2205	Математика (профильная)	27,20	27,82	< 0,001
		Математическая грамотность	7,02	7,08	0,172

Источник / Примечание: Данная таблица составлена авторами на основе анализа материалов

Рисунок 5

Прирост по результатам «Математики (профильная)» ЕНТ в разрезе типа школ

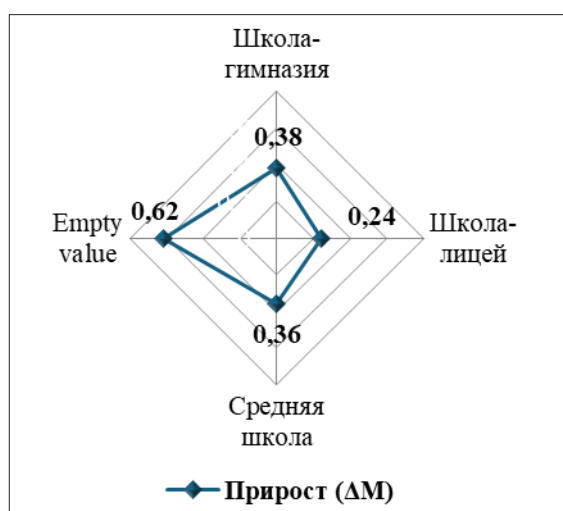
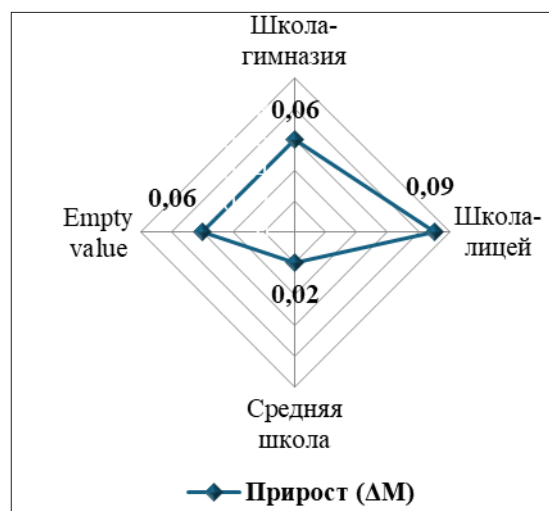


Рисунок 6

Прирост по результатам «Математической грамотности» ЕНТ в разрезе типа школ



Источник / Примечание: Данный рисунок составлен авторами на основе анализа материалов

Результаты ЕНТ показывают, что тип образовательного учреждения существенно влияет на достижения учащихся по математике и математической грамотности. Лицеи (тип 2) демонстрируют наивысшие средние показатели как по профильной математике (25,64 в первой попытке и 25,88 во второй), так и по математической грамотности (6,63 и 6,72 соответственно), а также являются единственным типом школ со статистически значимым прогрессом по обоим предметам между попытками (прирост $\Delta M = +0,24$, $p=0,014$ по математике; $\Delta M = +0,09$, $p=0,018$ по грамотности).

Гимназии (тип 1) показывают средние результаты (23,87 → 24,25 по математике; 6,36 → 6,42 по грамотности) со значимым приростом только по профильной математике ($\Delta M = +0,38$, $p<0,001$), в то время как улучшение по грамотности статистически не подтверждено ($\Delta M = +0,06$, $p=0,111$).

Обычные школы (тип 4) – наименьшие средние баллы (21,39 → 21,75 по математике; 5,96 → 5,98 по грамотности). Рост показателей математической грамотности в них статистически не значим ($\Delta M = +0,02$, $p=0,432$), что указывает на ограниченный эффект образова-

тельной среды в этих учреждениях для развития прикладных компетенций, несмотря на значимый прогресс по академической математике ($\Delta M = +0,36$, $p < 0,001$).

Динамика результатов подтверждает, что профильная математика (знания и алгоритмы) улучшается при подготовке почти во всех типах школ, а математическая грамотность (прикладные навыки) является более устойчивым конструктом, рост которого требует особых педагогических условий, эффективно создаваемых, по данным, в лицеях. Это подчеркивает необходимость учета типа учебного заведения при разработке программ поддержки (OECD, 2019).

Обсуждение результатов

Динамика баллов и эффект тестирования. Результаты исследования подтверждают гипотезу о том, что повторное участие в ЕНТ ведет к статистически значимому, хотя и умеренному росту баллов. Этот феномен объясняется «эффектом практики» – снижением ситуативной тревожности и повышением навыков прохождения тестирования (Baikenov & Najar, 2025). Однако дифференциация прироста по дисциплинам (высокий для профильной математики и низкий для математической грамотности) подчеркивает фундаментальные различия в когнитивных структурах этих предметов. Профильная математика в текущей конструкции теста больше опирается на алгоритмические знания, которые легче «подтянуть» за короткий срок между попытками. Математическая грамотность требует более глубокой переработки информации и навыков моделирования, которые формируются длительно и менее чувствительны к кратким интервенциям (Karpasova et al., 2025).

Психометрическая устойчивость как индикатор равенства. Выявленная высокая корреляция ($r = 0,88$) между попытками по математике свидетельствует о качественной разработке заданий и их сопоставимости. Однако стабильность результатов у разных групп учащихся трактуется по-разному. Отсутствие прироста баллов у сельских школьников и учащихся обычных школ по математической грамотности при наличии роста в городах и лицеях свидетельствует о «скрытом неравенстве» (Moldabekova et al., 2025). Городские учащиеся имеют доступ к более качественной обратной связи и ресурсам подготовки (курсы, репетиторы), что позволяет им эффективно использовать вторую попытку (Baikenov & Najar, 2025). В сельской местности первая по-

пытка часто остается единственным реальным опытом столкновения с форматом ЕНТ, но из-за отсутствия методической поддержки между попытками потенциал улучшения результата остается нереализованным (Nurbayev, 2021).

Территориальный детерминизм образовательных траекторий. Данные ANOVA и региональные срезы подтверждают наличие «географического потолка» достижений. Учащиеся южных и западных регионов (Туркестанская, Мангистауская области) находятся в зоне риска, демонстрируя баллы значительно ниже среднего национального уровня (Koreyeva, 2025). Это коррелирует с данными о социально-экономической депривации этих территорий, переуплотненности классов и дефиците ИКТ-инфраструктуры (Moldabekova et al., 2025). Особого внимания заслуживает Туркестанский регион: при самом большом количестве тестируемых его результаты остаются крайне низкими, что требует пересмотра стратегий регионального управления образованием (Koreyeva, 2025).

Напротив, успех Северо-Казахстанской области и г. Алматы обусловлен не только лучшим финансированием, но и сохранением сильных академических традиций преподавания математики. Высокая стабильность результатов в Алматы ($p = 0,353$) может указывать на то, что подготовка в этом мегаполисе настолько интенсивна, что учащиеся достигают своего истинного максимума уже в первом цикле.

Заключение

Психометрический анализ результатов ЕНТ-2025 по математическим дисциплинам позволяет сделать следующие выводы.

Во-первых, тест демонстрирует высокую надежность и сопоставимость результатов, особенно в блоке профильной математики ($r = 0,88$), что свидетельствует о качестве измерительного инструментария и соблюдении стандартов оценивания. Во-вторых, наличие второй попытки оказывается обоснованным для академических дисциплин, однако в краткосрочной перспективе она практически не влияет на показатели математической грамотности, особенно у социально уязвимых групп. В-третьих, сохраняется выраженный территориальный разрыв: регион и тип населенного пункта существенно определяют как абсолютные результаты, так и потенциал их улучшения, при этом сельские школы, а также западные и южные регионы демонстрируют устойчивое отставание. Наконец, выявлена ин-

ституциональная стратификация: лицеи и гимназии более эффективно используют возможности повторного тестирования по сравнению с общеобразовательными школами и колледжами.

В рамках будущих исследований предполагается рассмотрение вопросов, связанных с дифференциацией подходов к разработке измерительных материалов в зависимости от характера оцениваемого конструкта. В частности, планируется анализ возможностей усиления моделирующего и контекстного компонентов заданий в блоке «математическая грамотность», ориентированных на работу с нестандартными ситуациями и обоснованный выбор стратегии решения, с целью повышения содержательной валидности теста и формирования положительного обратного воздействия на образовательный процесс.

В отношении блока «профильная математика» предметом последующих исследований станет усиление психометрического контроля за эквивалентностью параллельных тестовых форм, используемых в различных периодах тестирования, с учетом выраженного эффекта практики и высокой надежности повторного тестирования. Это позволит обеспечить сопоставимость результатов и более корректную интерпретацию прироста баллов как показателя реального приращения знаний.

Отдельным направлением будущих исследований рассматривается институционализация регулярного анализа дифференциального функционирования заданий (DIF) по социо-территориальным признакам, включая регион и тип населенного пункта. Планируется проведение содержательной экспертизы заданий, демонстрирующих статистически значимый DIF, на предмет наличия конструктивно-нерелевантной сложности, с последующим использованием результатов для системного обновления банка тестовых заданий.

Кроме того, в будущих исследованиях предполагается использование эмпирических данных для целевого психометрического анализа зон устойчиво низких результатов, направленного на выявление проблемных тематических блоков и типов заданий. Полученные выводы

могут быть использованы для совершенствования подготовки разработчиков тестов и формирования более репрезентативных и валидных измерительных инструментов.

Финансирование

Работа выполнена в рамках инициативных исследований авторов на базе Национального центра тестирования.

Благодарность

Авторы выражают благодарность руководству Национального центра тестирования за предоставление деперсонализированных и агрегированных данных Единого национального тестирования 2025 года.

Заявление о конфликте интересов

Авторы заявляют о полном отсутствии конфликта интересов. Сбор, деперсонализация и статистическая обработка данных производились в строгом соответствии с принципами научной объективности, академической честности и независимости анализа.

Вклад авторов

Ш.Б. Алтыбаева – концептуализация исследования, формулировка методологической базы, научное руководство, критический анализ, подготовка текста разделов.

К.Т. Туралбаева – обзор научной литературы, анализ международного опыта применения классической теории тестов и IRT-моделирования.

Л.Е. Шинетова – сбор и верификация эмпирических данных, первичный дескриптивный анализ, расчет показателей надежности повторного тестирования.

Г.К. Жабеева – статистическая обработка результатов, проведение парных t-тестов, дисперсионного анализа, оформление табличного и графического материала.

Ж.Ж. Бейсенова – психометрический аудит, интерпретация результатов территориальных и институциональных различий, ведение научной переписки.

Литература

Министерство образования и науки Республики Казахстан. *Об утверждении Типовых правил проведения текущего контроля успеваемости, промежуточной и итоговой аттестации обучающихся для организаций образования, реализующих общеобразовательные учебные программы начального, основного среднего, общего среднего образования* (Приказ № 204 от 2 мая 2017 года и Приказ № 600 от 31 октября 2018 года). Астана: МОН РК, 2018.

Министерство образования и науки Республики Казахстан. *Об утверждении типовых учебных планов начального, основного среднего, общего среднего образования Республики Казахстан* (Приказ № 500 от 8 ноября 2012 года). – Астана: МОН РК, 2012.

Министерство просвещения Республики Казахстан. *Об утверждении государственных общеобязательных стандартов дошкольного воспитания и обучения, начального, основного среднего и общего среднего, технического и профессионального, послесреднего образования* (Приказ № 348 от 3 августа 2022 года). – Астана: МП РК, 2022.

Национальный центр тестирования. *Информация о структуре единого национального тестирования*. Астана, 2025. <https://testcenter.kz>

Abdrasilov, B., Charkhabi, M., Shinetova, L., Radi Afsouran, N., Kardanova, E. The Influence of Demographic Characteristics on Student Academic Performance in University Admission Tests // *Psychology in Russia: State of the Art*, 2025. -№18(4). P. 87–103. DOI: 10.11621/pir.2025.0405

AERA, APA, NCME. *Standards for educational and psychological testing* // American Educational Research Association. 2014.

Algina, J., Penfield, R. D. (2009). Classical test theory // In *The SAGE handbook of quantitative methods in psychology*. SAGE. 2009. P. 93–122. <https://doi.org/10.4135/9780857020994.n5>

Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., Zheng, Y. A validation review of the SAT and ACT for college and university admissions decisions // *Education Policy Analysis Archives*, 2025. №33(28). <https://doi.org/10.14507/epaa.33.8734>

Baikenov, K., Hajar, A. Buying your way into higher education? unpacking the association between fee-charging private tutoring and educational inequality in Kazakhstan // *Asia Pacific Education Review*. 2025. <https://doi.org/10.1007/s12564-025-10094-7>

Barry, A. E., Chaney, B. H., Piazza-Gardner, A. K., Chavarria, E. A. Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals // *Health Education & Behavior*, 2014. -№ 41(1). P. 12–18. DOI: 10.1177/1090198113483139

Bates, D., Maechler, M., Bolker, B., Walker, S. Fitting Linear Mixed-Effects Models Using lme4 // *Journal of Statistical Software*, 2015. №67(1). P. 1–48. DOI: 10.18637/jss.v067.i01

Bull, C., Byrnes, J., Hettiarachchi, R., Downes, M. (2019). A systematic review of the validity and reliability of patient-reported experience measures // *Health Services Research*, 2018. – №54(5). P. 1023–1035. DOI: 10.1111/1475-6773.13187

Chankseliani, M., Gorgodze, S., Janashia, S., Kurakbayev, K. Rural disadvantage in the context of centralised university admissions: A multiple case study of Georgia and Kazakhstan // *Comparative Education*, 2020. №56(7). P. 995–1013.

Didarbekova, N., Sarmurzin, Y., Altybaeva, S., Abdrasilov, B., Karkenova, A. Advancing digital item development in Asian assessment systems: Insights from Kazakhstan's Unified National Testing // *Frontiers in Education*, 2025. – №10. P. 1654674. <https://doi.org/10.3389/educ.2025.1654674>

Dorans, N. J. Two new approaches to assessing differential item functioning: Standardization and the Mantel–Haenszel method // *Applied Measurement in Education*, 1989. – №2(3). P.217–233. https://doi.org/10.1207/s15324818ame0203_3

Egov.kz. SNT in Kazakhstan: preparation and order of conduction | Electronic government of the Republic of Kazakhstan. Астана, 2024. https://egov.kz/cms/en/articles/about_ent

Jumabayeva, Z. The key drivers of the Unified National Test in Kazakhstan: A critical analysis of its impact on school leavers // *NUGSE Research in Education*, 2026. – № 1(2). – P.16–20. Retrieved from nugserie.nu.edu.kz

Kappassova, S., Abylkassymova, A., Bulut, U., Zykrina, S., Zhumagulova, Z., Balta, N. Mathematical literacy and its influencing factors: A decade of research findings (2015–2024) // *Eurasia Journal of Mathematics, Science and Technology Education*, 2025. -№ 21(7). P. em2671. <https://doi.org/10.29333/ejmste/16615>

Kopeyeva, A. *Beyond the scores: A critical realist study of the causes of regional educational inequality in Kazakhstan* [Apollo – University of Cambridge Repository]. 2025. <https://doi.org/10.17863/CAM.119294>

Marini, J. P., Westrick, P. A., Young, L., Shaw, E. J. *Validity of the SAT® for enrollment-related decisions*. College Board. 2020. <https://reports.collegeboard.org/media/pdf/validity-sat-enrollment-related-decisions.pdf>

Messick, S. The interplay of evidence and consequences in the validation of performance assessments // *Educational Researcher*, 1994. №23(2). P.13–23. <https://doi.org/10.3102/0013189X023002013>

Messick, S. Validity and Washback in Language Testing // *ETS Research Report Series*, 1996(1). – i–18. <https://doi.org/10.1002/j.2333-8504.1996.tb01695.x>

Mingisheva, N. Development and Challenges of Standardized Testing in Kazakhstan: Transition from National to International Standards // *Journal of Educational Sciences*, 2023. №76(3). – P. 94–103. <https://doi.org/10.26577/JES.2023.v76.i3.08>

Moldabekova, A., Satybaldin, A., Omir, A., Sadykov, E., Beimisheva, A. Household deprivation in Kazakhstan: A factor analysis of regional disparities // *Societies*, 2025. №15(9). P. 254. <https://doi.org/10.3390/soc15090254>

Nurbaev, Z. *Inequality Between Students of Rural and Urban Schools in Kazakhstan: Causes and Ways to Address It*. Central Asia Program. 2021. <https://centralasiaprogram.org/publications-all/research-updates/fellows-papers/inequality-between-students-of-rural-and-urban-schools-in-kazakhstan-causes-and-ways-to-address-it/>

OECD. *PISA 2022 Results: Country Note – Kazakhstan*. OECD Publishing, 2023. <https://www.oecd.org/en/publications/pisa-2022-results-kazakhstan.html>

OECD. *Strengthening national examinations in Kazakhstan to achieve national goals*. OECD Publishing, 2020.

Sanchez, E. I. *Predicting STEM achievement: A comparative study of ACT® scores and high school GPA* (ACT Research Report No. R2501). ACT Education, 2025.

Smith, E. V., Jr. Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective // *Journal of Applied Measurement*, 2021. – №2(3). – P. 281–311.

Woodruff, D., Wu, Y. F. *Statistical considerations in choosing a test reliability coefficient* (ACT Research Report No. 2012-10). ACT, Inc, 2012.

References

Abdrasilov, B., Charkhabi, M., Shinetova, L., Radi Afsouran, N., & Kardanova, E. (2025). The Influence of Demographic Characteristics on Student Academic Performance in University Admission Tests. *Psychology in Russia: State of the Art*, 18(4), 87–103. DOI: 10.11621/pir.2025.0405

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Algina, J., & Penfield, R. D. (2009). Classical test theory. In *The SAGE handbook of quantitative methods in psychology* (pp. 93–122). SAGE. <https://doi.org/10.4135/9780857020994.n5>

Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarov, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28). <https://doi.org/10.1407/epaa.33.8734>

Baikenov, K., & Hajar, A. (2025). Buying your way into higher education? unpacking the association between fee-charging private tutoring and educational inequality in Kazakhstan. *Asia Pacific Education Review*. <https://doi.org/10.1007/s12564-025-10094-7>

Barry, A. E., Chaney, B. H., Piazza-Gardner, A. K., & Chavarria, E. A. (2014). Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals. *Health Education & Behavior*, 41(1), 12–18. DOI: 10.1177/1090198113483139

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: 10.18637/jss.v067.i01

Bull, C., Byrnes, J., Hettiarachchi, R., & Downes, M. (2019). A systematic review of the validity and reliability of patient-reported experience measures. *Health Services Research*, 54(5), 1023–1035. DOI: 10.1111/1475-6773.13187

Chankseliani, M., Gorgodze, S., Janashia, S., & Kurakbayev, K. (2020). Rural disadvantage in the context of centralised university admissions: A multiple case study of Georgia and Kazakhstan. *Comparative Education*, 56(7), 995–1013.

Didarbekova, N., Sarmurzin, Y., Altybaeva, S., Abdrasilov, B., & Karkenova, A. (2025). Advancing digital item development in Asian assessment systems: Insights from Kazakhstan's Unified National Testing. *Frontiers in Education*, 10, 1654674. <https://doi.org/10.3389/educ.2025.1654674>

Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel–Haenszel method. *Applied Measurement in Education*, 2(3), 217–233. https://doi.org/10.1207/s15324818ame0203_3

egov.kz. (2024). *SNT in Kazakhstan: preparation and order of conduction* | *Electronic government of the Republic of Kazakhstan*. https://egov.kz/cms/en/articles/about_ent

Jumabayeva, Z. (2016). The key drivers of the Unified National Test in Kazakhstan: A critical analysis of its impact on school leavers. *NUGSE Research in Education*, 1(2), 16–20.

Kappassova, S., Abylkassymova, A., Bulut, U., Zykrina, S., Zhumagulova, Z., & Balta, N. (2025). Mathematical literacy and its influencing factors: A decade of research findings (2015–2024). *Eurasia Journal of Mathematics, Science and Technology Education*, 21(7), em2671. <https://doi.org/10.29333/ejmste/16615>

Kopeyeva, A. (2025). *Beyond the scores: A critical realist study of the causes of regional educational inequality in Kazakhstan* [Doctoral dissertation, Apollo – University of Cambridge Repository]. <https://doi.org/10.17863/CAM.119294>

Marini, J. P., Westrick, P. A., Young, L., & Shaw, E. J. (2020). *Validity of the SAT® for enrollment-related decisions*. College Board.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23. <https://doi.org/10.3102/0013189X023002013>

Messick, S. (1996). Validity and Washback in Language Testing. *ETS Research Report Series*, 1996(1), i–18. <https://doi.org/10.1002/j.2333-8504.1996.tb01695.x>

Mingisheva, N. (2023). Development and Challenges of Standardized Testing in Kazakhstan: Transition from National to International Standards. *Journal of Educational Sciences*, 76(3), 94–103. <https://doi.org/10.26577/JES.2023.v76.i3.08>

Ministerstvo obrazovaniya i nauki Respubliki Kazakhstan. (2012). *Ob utverzhdenii tipovykh uchebnykh planov nachal'nogo, osnovnogo srednego, obshchego srednego obrazovaniya Respubliki Kazakhstan* [On approval of model curricula for primary, basic secondary, general secondary education of the Republic of Kazakhstan] (Order No. 500 of November 8, 2012).

Ministerstvo obrazovaniya i nauki Respubliki Kazakhstan. (2017). *Ob utverzhdenii Tipovykh pravil provedeniya tekushchego kontrolya uspevaemosti, promezhutochnoy i itogovoy attestatsii obuchayushchikhsya dlya organizatsiy obrazovaniya, realizuyushchikh obshcheobrazovatel'nye uchebnye programmy nachal'nogo, osnovnogo srednego, obshchego srednego obrazovaniya* [On approval of the Model rules for ongoing monitoring of academic performance, intermediate and final certification of students for educational organizations implementing general education curricula of primary, basic secondary, general secondary education] (Order No. 204 of May 2, 2017 and Order No. 600 of October 31, 2018).

Ministerstvo prosveshcheniya Respubliki Kazakhstan. (2022). *Ob utverzhdenii gosudarstvennykh obshcheobrazovatel'nykh standartov doshkol'nogo vospitaniya i obucheniya, nachal'nogo, osnovnogo srednego i obshchego srednego, tekhnicheskogo i*

professional'nogo, poslesrednego obrazovaniya [On approval of state compulsory standards for preschool upbringing and training, primary, basic secondary and general secondary, technical and vocational, post-secondary education] (Order No. 348 of August 3, 2022).

Moldabekova, A., Satybaldin, A., Omir, A., Sadykov, E., & Beimisheva, A. (2025). Household deprivation in Kazakhstan: A factor analysis of regional disparities. *Societies*, 15(9), 254. <https://doi.org/10.3390/soc15090254>

Natsional'nyy tsentr testirovaniya. (2025). *Informatsiya o strukture edinogo natsional'nogo testirovaniya* [Information on the structure of the unified national testing]. <https://testcenter.kz>

Nurbaev, Z. (2021). *Inequality Between Students of Rural and Urban Schools in Kazakhstan: Causes and Ways to Address It*. Central Asia Program. <https://centralasiaprogram.org/publications-all/research-updates/fellows-papers/inequality-between-students-of-rural-and-urban-schools-in-kazakhstan-causes-and-ways-to-address-it/>

OECD. (2020). *Strengthening national examinations in Kazakhstan to achieve national goals*. OECD Publishing.

OECD. (2023). *PISA 2022 Results: Country Note – Kazakhstan*. OECD Publishing. <https://www.oecd.org/en/publications/pisa-2022-results-kazakhstan.html>

Sanchez, E. I. (2025). *Predicting STEM achievement: A comparative study of ACT® scores and high school GPA* (ACT Research Report No. R2501). ACT, Inc.

Smith, E. V., Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2(3), 281–311.

Woodruff, D., & Wu, Y. F. (2012). *Statistical considerations in choosing a test reliability coefficient* (ACT Research Report No. 2012-10). ACT, Inc.

Сведения об авторах:

Шинетова Ляззат Ермековна – магистр, Национальный центр тестирования, заведующий лабораторией (Астана, Казахстан, e-mail: shinetovalyazzat24@gmail.com)

Жабаева Гульдана Куатовна – магистр, Национальный центр тестирования, ведущий научный сотрудник (Астана, Казахстан, e-mail: danka1919@mail.ru)

Бейсенова Жанар Жумагазиновна – магистр, Национальный центр тестирования, ведущий научный сотрудник (Астана, Казахстан, e-mail: zhanar-beisenova@mail.ru)

Адулова Елена Владимировна – магистр, «Центр педагогических измерений» АОО Назарбаев Интеллектуальные школы (Астана, Казахстан, e-mail: lenarakhmail@gmail.com)

Есеналиева Салтанат Сериковна – обладатель степени магистра Brunel University London, заместитель директора Департамента учебно-методической работы АО «Национальный центр повышения квалификации «Өрлеу» (Астана, Казахстан, e-mail: yessenaliyevass@gmail.com)

Information about authors:

Shinetova Lyazzat Ermekovna – Master's degree, National Testing Center, Head of the laboratory (Astana, Kazakhstan, e-mail: shinetovalyazzat24@gmail.com)

Zhabayeva Guldana Kuatovna – Master's degree, National Testing Center, Leading Researcher (Astana, Kazakhstan, e-mail: danka1919@mail.ru)

Beisenova Zhanar Zhumagazinovna – Master's degree, National Testing Center, Leading Researcher (Astana, Kazakhstan, e-mail: zhanar-beisenova@mail.ru)

Adulova Yelena – master of science in educational leadership, «Center for Pedagogical Measurements» under AEO «Nazarbayev Intellectual Schools» (Astana, Kazakhstan, e-mail: lenarakhmail@gmail.com)

Yessenaliyeva Saltanat – MA, Brunel University London, Deputy Director of the Department of Educational and Methodological Work, JSC National Center for Professional Development «Orleu» (Astana, Kazakhstan, e-mail: yessenaliyevass@gmail.com)

Авторлар туралы мәлімет:

Шинетова Ляззат Ермековна – магистр, Ұлттық тестілеу орталығы, зертхана меңгеруші (Астана, Қазақстан, e-mail: shinetovalyazzat24@gmail.com)

Жабаева Гульдана Куатовна – магистр, Ұлттық тестілеу орталығы, жетекші ғылыми қызметкер (Астана, Қазақстан, e-mail: danka1919@mail.ru)

Бейсенова Жанар Жумагазиновна – магистр, Ұлттық тестілеу орталығы, жетекші ғылыми қызметкер (Астана, Қазақстан, e-mail: zhanar-beisenova@mail.ru)

Адулова Елена Владимировна – магистр, «Педагогикалық өлшеулер орталығы» Назарбаев Зияткерлік мектептері ДББҰ (Астана, Қазақстан, e-mail: lenarakhmail@gmail.com)

Есеналиева Салтанат Сериковна – Brunel University London университетінің магистрі, «Өрлеу» біліктілікті арттыру ұлттық орталығы» АҚ Оқу-әдістемелік жұмыс департаменті директорының орынбасары (Астана, Қазақстан, e-mail: yessenaliyevass@gmail.com)

Поступило: 13.01.2026

Принято: 25.05.2026